REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY) 05-01-2016	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 28-Aug-2012 - 27-Aug-2013

4. TITLE AND SUBTITLE
Final Report: A Computing Platform for Parallel Sparse Matrix Computations

5a. CONTRACT NUMBER
W911NF-12-1-0447

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611103

6. AUTHORS
Ahmed H. Sameh, Alicia Klinvex, Yao Zhu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Purdue University
155 S. Grant Street

West Lafayette, IN 47907 -2114

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
61443-CS-RIP.5

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
This grant enabled the purchase of an Intel multiprocessor consisting of eight multicore nodes interconnected via an infiniband. Each node contains 24 cores. This parallel computing platform has been used by my research group in the early stages of developing large sparse linear system and symmetric eigenvalue problem solvers (ARO grant W911NF-07-R-0003-04) that are suitable for parallel architectures containing hundreds of multicore nodes (thousands of cores). Once our parallel solvers obtain the correct solutions, and perform properly on this 8-node platform, the codes are ported to much larger parallel computing platforms at the Intel Corporation. Two Ph.D

15. SUBJECT TERMS
large sparse linear system solvers, large sparse symmetric eigenvalue problem solvers, parallel programming

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Ahmed Sameh
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU	UU		19b. TELEPHONE NUMBER 765-494-1559

Report Title

Final Report: A Computing Platform for Parallel Sparse Matrix Computations

ABSTRACT

This grant enabled the purchase of an Intel multiprocessor consisting of eight multicore nodes interconnected via an infiniband. Each node contains 24 cores. This parallel computing platform has been used by my research group in the early stages of developing large sparse linear system and symmetric eigenvalue problem solvers (ARO grant W911NF-07-R-0003-04) that are suitable for parallel architectures containing hundreds of multicore nodes (thousands of cores). Once our parallel solvers obtain the correct solutions, and perform properly on this 8-node platform, the codes are ported to much larger parallel computing platforms at the Intel Corporation. Two Ph.D. dissertations have resulted from the above grant (awarded by Dr. Myers), and this DURIP grant (awarded by Dr. Coyle).

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 3.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper
01/05/2016	2.00	Yao Zhu, Ahmed H. Sameh. SPIKE+: A family of parallel hybrid sparse linear system solvers, Journal of Computational and Applied Mathematics (05 2015)

TOTAL: **1**

Number of Manuscripts:

Books

Received	Book

TOTAL:

Received		Book Chapter
01/05/2016	1.00	Yao Zhu, Ahmed H. Sameh. How to generate effective block Jacobi preconditioners for solving large sparse linear systems, Yet to be published: Birkhauser, (02 2016)

TOTAL: **1**

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Yao Zhu	0.50	
Alicia Klinvex	0.10	
FTE Equivalent:	**0.60**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Ahmed H. Sameh	0.10	
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME	
Yao Zhu	
Alicia Klinvex	
Total Number:	**2**

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Two classes of parallel solvers have been developed. The first is a family of parallel sparse linear system solvers -- the PSPIKE family (Spike/Pardiso hybrid solvers). The second is a family of Trace Minimization eigensolvers for the symmetric standard and generalized eigenvalue problems. Both classes of solvers proved to be competitive with existing solvers in being more robust and more scalable on parallel architectures with superior speeds. In particular the TraceMin eigensolvers developed through these two ARO grants have been adopted by the DOE Trilinos project at Sandia National Laboratories.

Technology Transfer

TraceMIN: a scalable parallel algorithm for large sparse symmetric eigenvalue problems

Alicia Klinvex

Sandia National Labs.

Ahmed H. Sameh

Computer Science, Purdue University

Support: ARO, Intel, NSF

St Girons: June 27 – July 2; 2015

Symmetric Generalized Eigenvalue Problem

$$Ax = \lambda Bx$$

A is sparse and symmetric

B is sparse and s.p.d.

Obtain the p eigenvalues closest to 0 and their corresponding eigenvectors, $p << n$

The Trace minimization scheme

$$\min_{Y^T BY = I_p} tr(Y^T AY) = \sum_{i=1}^{p} \lambda_i$$

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p < \lambda_{p+1} \leq \cdots \leq \lambda_n$$

$$Y \in R^{n\times p} \quad ; \quad p << n.$$

A.S. & J. Wisniewski: SINUM, 1982

A.S. & Z. Tong: J. Comp. Appl. Math., 2000.

Extensions of this optimization problem for definite matrix pair

- *B: indefinite nonsingular*
  - *Kobac-Striko and K. Veselic (1995)*
- *B: indefinite singular*
  - *Liang, Li, and Z. Bai*

A and B are called "definite pair" if there exists a scalar μ such that (A – μ B) is s.p.d , or s.p.s.d *solve*

$$\begin{pmatrix} A & BY_k \\ Y_k^T B & O \end{pmatrix} \begin{pmatrix} \Delta_k \\ L_k \end{pmatrix} = \begin{pmatrix} AY_k \\ O \end{pmatrix} \qquad *$$

or

$$\begin{pmatrix} A & BY_k \\ Y_k^T B & O \end{pmatrix} \begin{pmatrix} Y_k - \Delta_k \\ -L_k \end{pmatrix} = \begin{pmatrix} O \\ I_p \end{pmatrix} \qquad **$$

- *(Y_k – Δ_k) is used to generate Y_{k+1}*
- *tr (Y_{k+1} A Y_{k+1}) ≤ tr(Y_k A Y_k)*
- *Different schemes (direct/iterative) for solving (*) or (**)*
- *TraceMIN does not require low relative residuals in solving (*) or (**)*

a) *Using a direct sparse linear system solver (Pardiso)*

- *Solve*

  $AZ_k = BY_k$ *for* Z_k

- *Solve*

$$\begin{pmatrix} I & Z_k \\ Y_k^T B & O \end{pmatrix} \begin{pmatrix} Y_k - \Delta_k \\ -L \end{pmatrix} = \begin{pmatrix} O \\ I_p \end{pmatrix}$$

for $(Y_k - \Delta_k)$

b) Use a Krylov subspace method to solve

$$(I - P_k)A(I - P_k)\Delta_k = (I - P_k)AY_k$$

for Δ_k

where

$$P_k = BY_k(Y_k^T B^2 Y_k)^{-1} Y_k^T B.$$

e.g. CG, or Minres; rel. res ≤ 10^{-6}

c) Solve

$$\begin{pmatrix} A & BY \\ Y^T B & O \end{pmatrix} \begin{pmatrix} Y - \Delta \\ -L \end{pmatrix} = \begin{pmatrix} O \\ I_p \end{pmatrix}$$

via a Krylov subspace method with the preconditioner

$$M = diag\,(\hat{A}\,,\, Y^T B\, \hat{A}^{-1} B\, Y)$$

Â is an approximation of A

Murphy, Golub, Wathen: SISC 1999

TraceMIN algorithm

- Choose initial (n x s) block V; $s \approx 2p$
- <u>Repeat until convergence</u>
  - 1. B-orthonormalize V
  - 2. Form approximate eigenpairs

    $Y^T A Y = \Sigma, \quad Y^T B Y = I; \; \Sigma = diag(\sigma_1, \sigma_2, \ldots., \sigma_s)$
  - 3. $R = AY - BY\Sigma$, check $norm(r_k) / \sigma_k \leq 10^{-5}$
  - 4. Solve saddle-point problem for Δ or $Y - \Delta$
  - 5. update $V = Y - \Delta$

Convergence

- Convergence bounded by $| \lambda_k / \lambda_{s+1} |$
- Origin shifts can lead to faster convergence
  - $(A - \mu_k B)\, x_k = (\lambda_k - \mu_k)\, x_k$
  - $| (\lambda_k - \mu_k) / \lambda_{s+1} |$
- If A is s.p.d, then

$$\Phi_k = (y_k - x_k)^T A\, (y_k - x_k)$$

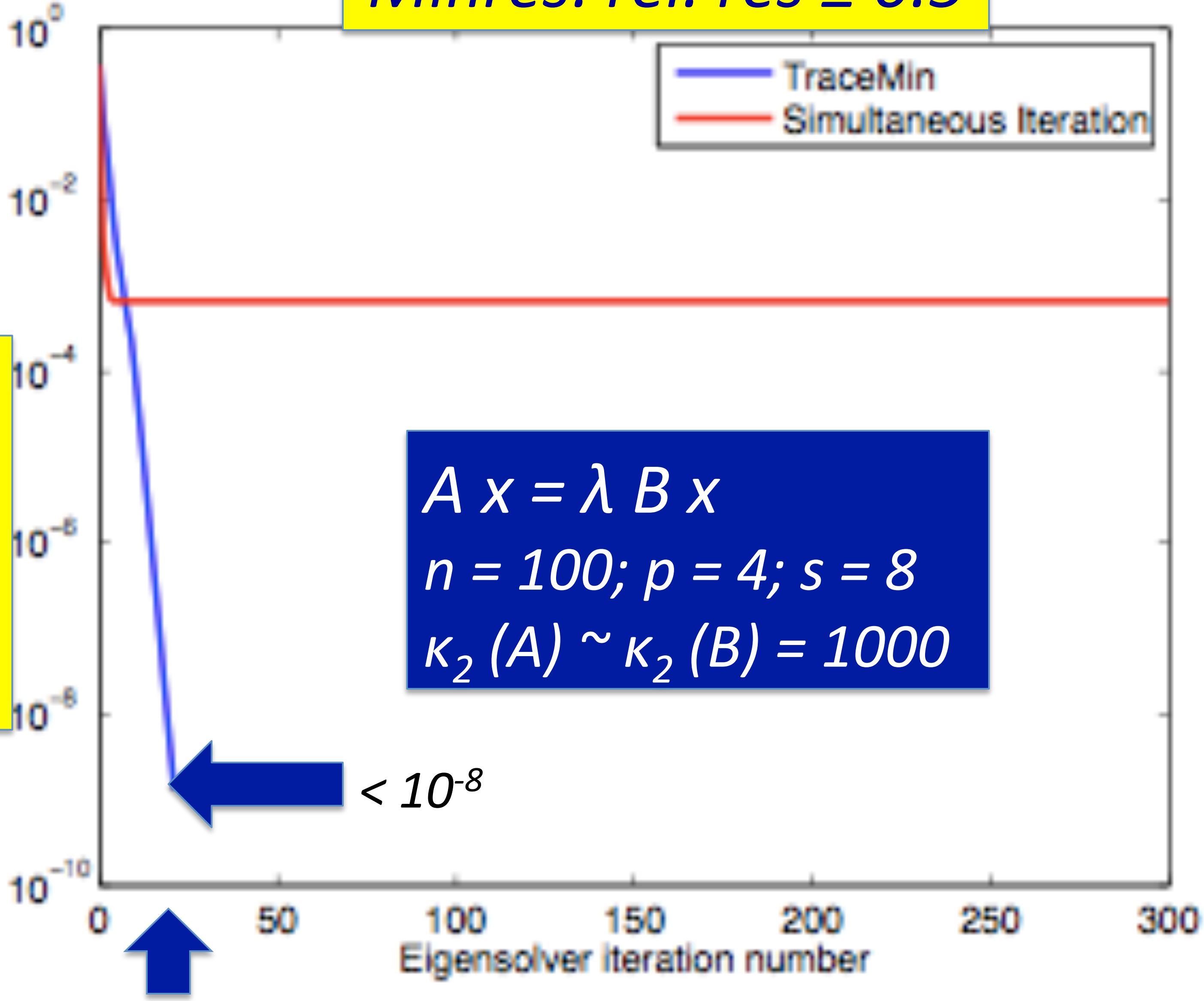
is reduced asymptotically by a factor of

$(\lambda_k / \lambda_{s+1})^2$

Simultaneous Iterations & TraceMIN

- *Simultaneous Iterations is mathematically equivalent to TraceMIN.*
- *In each step of Simultaneous Iterations, we need to solve systems of the form $A\,V = B\,Y$.*
- *If we solve such systems, as well as TraceMIN saddle-point problems using Minres, for example, with a modest stopping criterion, i.e. rel. res. ≤ 0.5, we get vastly different behavior in reduction of the trace.*

Minres: rel. res ≤ 0.5



Error in trace

$A\,x = \lambda\, B\, x$

$n = 100;\ p = 4;\ s = 8$

$\kappa_2\,(A) \sim \kappa_2\,(B) = 1000$

$< 10^{-8}$

< 25 *iterations*

Three trace minimization schemes for the symmetric eigenvalue problem

Solver	*TraceMIN 1982*	*Jacobi-Davidson (JD – 1996)*	*TraceMIN-Davidson (TD-2000)*
Subspace dimension	*constant*	*expanding*	*expanding*
Choice of shifts	*conservative*	*aggressive*	*conservative*

How to handle expanding subspaces

V := [V, Δ]

Upon restart, retain only relevant part of the subspace

Performance Results: a sample

Codes in Trilinos framework

TraceMIN & TraceMIN-Davidson (TD) vs:

1. The DOE Trilinos Anasazi package –

- *Krylov Schur (KS) [G.W. Stewart]; (BKS) [Saad]*
- *Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG), [Knyazev]*
- *Riemannian Trust Region (RTR) ** [Baker et al]*

*2. The SLEPc package -- Jacobi-Davidson (JD) ***

*** TraceMIN-based*

Robustness

- *Obtain the 4 smallest eigenpairs of the standard eigenvalue problem:* $A\,x = \lambda\, x$
- *Source: Tim Davis sparse matrix collection*
- *3 benchmarks for which A is symmetric positive definite*
- *1 benchmark for which A is indefinite*
- ✔*: success*
- ✗*: failure (no convergence or wrong multiplicity)*

Computing Platform

- *Intel cluster: Endeavor*
- *Infiniband interconnection network*
- *Each node consists of two 14-core Intel processors running at 2.6 GHz*
- *64 GB per node*
- *Hybrid programming paradigm MPI/OpenMP*

Robustness

Matrix	Type n	TD	BKS	LOBPCG	RTR	KS	JD
Poisson	spd 64 M	✔	✔	✗	✔	✗	✗
Flan_1565	spd 1.5 M	✔	✗	✗	✔	✗	✔
Hook_1498	spd 1.5 M	✔	✗	✗	✔	✗	✔
nlpkkt240	Indef. 28 M	✔	✗	✗	✗	✗	✔
		ANASAZI				SLEPc	

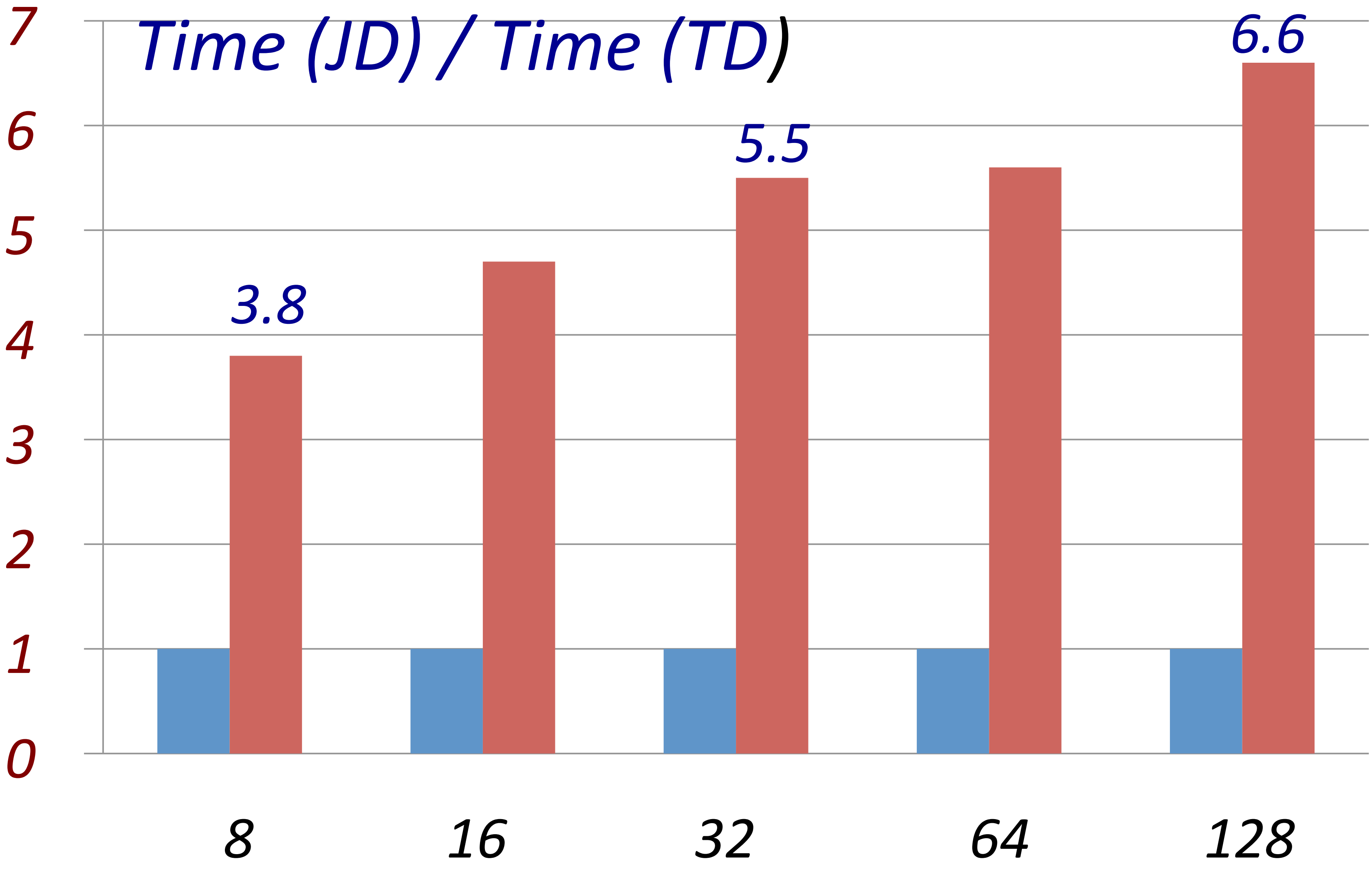
Time(algorithm X) / Time(TD)

- *32 nodes (24 cores per node)*

matrix	*Type*	*n*	*TD*	*BKS*	*LOBPCG*	*RTR*	*KS*	*JD*
Hook_1498	*spd*	*1.5 M*	*1.0*	✗	✗	*3.3*	✗	*1.5*

- *128 nodes (24 cores per node)*

matrix	*Type*	*n*	*TD*	*BKS*	*LOBPCG*	*RTR*	*KS*	*JD*
Poisson	*spd*	*64 M*	*1.0*	*19.7*	*3*	*1.6*	✗	*--*
Flan_1565	*spd*	*1.5 M*	*1.0*	✗	✗	*1.3*	✗	*2.1*
nlpkkt240	*indef*	*28 M*	*1.0*	✗	✗	✗	✗	*6.6*



computing the 4 smallest eigenpairs of nlpkkt240

TraceMIN: multisectioning & sampling (codes in Fortran 90)

- *Multisectioning:*
  - *For obtaining a large number of interior eigenvalues and their corresponding eigenvectors*
- *Sampling:*
  - *Obtaining several eigenpairs in the neighborhood of a given number of points within the spectrum*

Multisectioning

- *Using several shifts α_j we can obtain that interval (0,τ) that contains the number of eigenpairs desired.*
- *Divide the interval under consideration into many subintervals, one subinterval per node; each containing roughly the same number of eigenvalues (e.g. 20), if possible*
- *For each node j, we use our shared-memory TraceMIN eigensolver to obtain the smallest eigenpairs of $(A - \alpha_j B)\, x = (\lambda - \alpha_j)\, B\, x$.*

With the shifts α, we get saddle-point problems of the form:

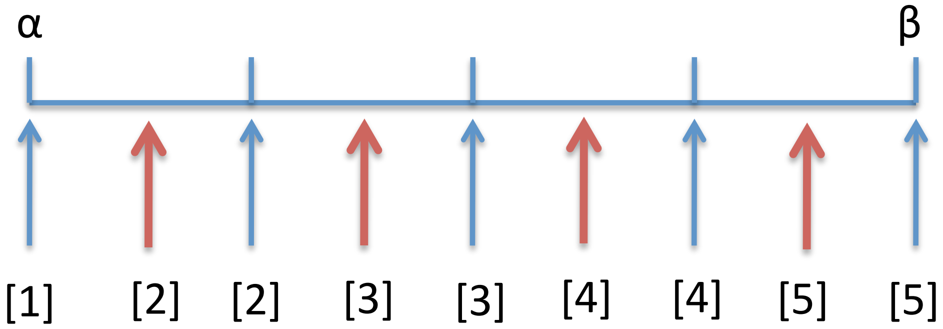
$$\begin{pmatrix} A-\alpha B & BY_k \\ Y_k^T B & O \end{pmatrix} \begin{pmatrix} Y_k - \Delta_k \\ -L_k \end{pmatrix} = \begin{pmatrix} O \\ I_p \end{pmatrix}.$$

A sparse direct solver (Pardiso) is used to Obtain the factorization of the indefinite (1,1) block:

$$A-\alpha B = LDL^T .$$

Multisectioning – *an illustration*

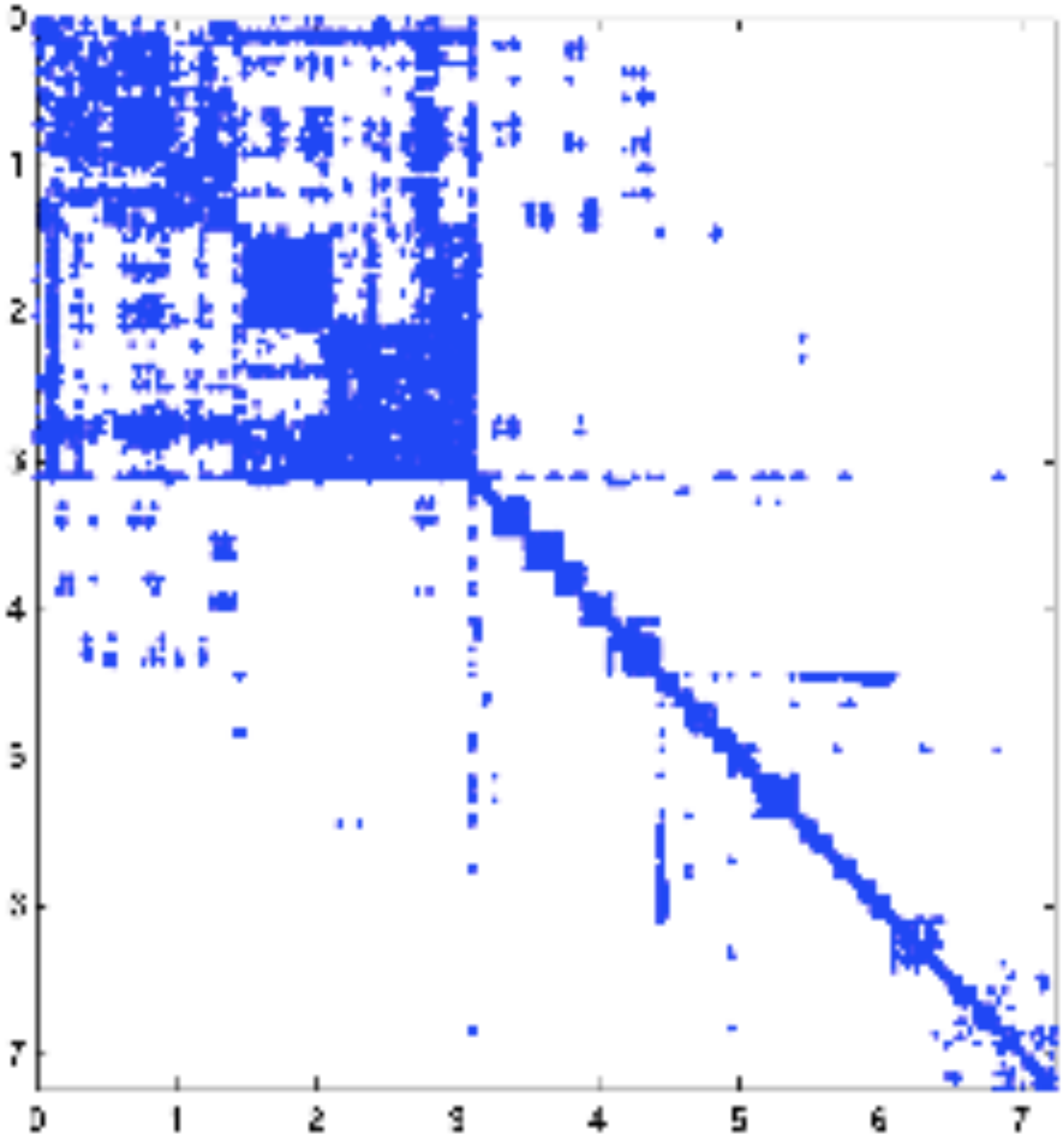
- *Obtain the eigenpairs corresponding to* [α, β]
- [α, β] is divided into 4 sub-intervals
- *Use five nodes*
- *Except for Node #1, each node performs 2 factorizations:*
  - *one for computing the inertia (blue), and*
  - *one for the TraceMIN iteration (red).*
- *Node 1 performs one factorization with shift α*



A NASTRAN benchmark

Car body dynamics –

- –*A x = λ B x (order: 1.5 M, and 7.2 M)*
- –*$A = A^T$, B := s.p.d.*
- – *norm(R,inf) ≤ 10^{-6}*
- –*Compute the smallest 1000 eigenpairs using 129 nodes.*



Sparsity Structure of A

n ~ 7.2 M

TraceMIN w/ multisectioning
vs. Intel's Math Kernel Library's FEAST

Problem	size	# of smallest λs	TraceMIN (seconds)	FEAST (seconds)	Speed Improve.
Nastran 1	1.5 M	1000	59	121	2.2
Nastran 2	7.2 M	1000	418	Failure	

All Trilinos eigensolvers failed on this NASTRAN benchmark

TraceMIN w/ multisectioning
vs. Intel's Math Kernel Library's FEAST…
using 129 nodes; norm $(r_k\ , inf)/\sigma_k \leq 10^{-5}$

Problem	size	# of λs interval	TraceMIN (seconds)	FEAST (seconds)	Speed Improve.
Anderson*	1.0 M	1143 [-.01,.01]	792	7910	10.0
af_shell10	1.5 M	1045 [2000,2250]	37	274	7.4

•Anderson Model of Localization (Schrödinger eq.): H x = λ x
diag (H): random; offdiag(H) = 1

Sampling: computing 400 eigenpairs

Matrix	*Size*	*Interval*	*abs. res.* *10 $^{-5}$*	*abs. res.* *10 $^{-9}$*
NASTRAN 1	*1.5 M*	[-.01, 10^3]	*21 s* *~.05 s/ep*	*40 s* *~ .10 s/ep*
NASTRAN 2	*7.2 M*	[-.01, 3 * 10^6]	*181 s* *~.45 s/ep*	*302 s* *~.76 s/ep*

Obtain the 4 eigenvalues closest to 100 points uniformly distributed in each interval and the corresponding eigenvectors using TraceMIN *on 100 nodes*

Thank you!